



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : G06F 15/403	A2	(11) International Publication Number: WO 91/06059 (43) International Publication Date: 2 May 1991 (02.05.91)
<p>(21) International Application Number: PCT/GB90/01554</p> <p>(22) International Filing Date: 9 October 1990 (09.10.90)</p> <p>(30) Priority data: 8922755.7 10 October 1989 (10.10.89) GB</p> <p>(71) Applicant (for all designated States except CA US): AMBIT RESEARCH LIMITED [GB/GB]; London House, 100 Kings Road, London SW6 4LX (GB).</p> <p>(71)(72) Applicant and Inventor (for CA US only): CAWLEY, William, Frederick [GB/GB]; Bircher Hall, Leominster, Herefordshire HR6 0AX (GB).</p> <p>(74) Agent: SHAW, Laurence; George House, George Road, Edgbaston, Birmingham B15 1PG (GB).</p>		<p>(81) Designated States: AT, AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH, CH (European patent), CM (OAPI patent), DE, DE (European patent), DK, DK (European patent), ES, ES (European patent), FI, FR (European patent), GA (OAPI patent), GB, GB (European patent), GR, GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU, LU (European patent), MC, MG, ML (OAPI patent), MR (OAPI patent), MW, NL, NL (European patent), NO, RO, SD, SE, SE (European patent), SN (OAPI patent), SU, TD (OAPI patent), TG (OAPI patent), US.</p> <p>Published Without international search report and to be republished upon receipt of that report.</p>
<p>(54) Title: DATA MANIPULATION</p> <p>(57) Abstract</p> <p>A spreadsheet layout is set up on a visual display unit of a computer having a database in which each item of data is stored in a separate record which includes a label in addition to the data itself, by the steps of: (a) naming a new worksheet and presenting it blank; (b) naming and positioning concepts as row headings and column headings in cells of the worksheet; (c) saving the worksheet in a usage file; (d) inspecting the layout to determine which cells should contain data to be obtained from the database and to determine the labels of any data which should be contained in each cell; and (e) searching for matches between the labels so determined and labels of data items in the database and, if found, displaying those items in the relevant cells of the worksheet.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MR	Mauritania
BE	Belgium	GA	Gabon	MW	Malawi
BF	Burkina Faso	GB	United Kingdom	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	SD	Sudan
CF	Central African Republic	KP	Democratic People's Republic of Korea	SE	Sweden
CG	Congo	KR	Republic of Korea	SN	Senegal
CH	Switzerland	LI	Liechtenstein	SU	Soviet Union
CI	Côte d'Ivoire	LK	Sri Lanka	TD	Chad
CM	Cameroon	LU	Luxembourg	TC	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

DATA MANIPULATION

The invention relates to the manipulation of data, and
5 more particularly to such manipulation using computer and
like equipment. More especially, the invention relates to
the presentation of data on the visual display unit of say a
personal computer in the manner of an electronic spreadsheet,
i.e. a layout analogous to an accountant's ledger sheet
10 having many columns and row in which data are entered. The
data may then be subjected to a variety of management tasks
such as audit, forecast, variable analysis, market planning,
corporate strategy planning and the like.

15 The advent of the personal computer has highlighted the
need for suitable programs whereby managers can store and
manipulate their own information without calling on the
skills of outside technicians to construct programs for them.
The spreadsheet presents the user with a cell structure on a
20 two dimensional plane, the cells of which may be defined by
the user to be constant or to be based on formulae, which are
enacted as the source values of the formulae are entered or
changed on demand.

25 The conventional spreadsheet stores its information
within the same structure as the cell headings, and each data
value has no significance outside its position in the
spreadsheet, thus limiting the amount of data that may be
handled with confidence to a very small proportion of that
30 which may be required.

In this specification the following terms have the
respective meanings unless the context otherwise requires

35 **WORKSHEET** is a spreadsheet tabular layout of columns and
rows adapted to do one or more specific tasks.

- CONCEPT is a name defined by the user to have a meaning within the Work sheet and which may be the name of the Worksheet or names in the row headings or column headings or the like.
- 5 ENTITY is a concept having a value (numeric, alphabetic or both) defined by a number of "attributes" (which are themselves concepts). For example, "Total Sales" may be an entity defined by attributes of Product, Region, and Period. All values of Total Sales would require Product, Region and Period to be specified, e.g. Total Sales/socks/UK/1989.
- 10 CELL is the smallest object in a Worksheet table the location of which is defined by a column position and a row position and which may contain numerical values or text. The information in a cell is defined by the Entities defining the column headings and row headings and the context (see below). For example those cells that may hold data are usually those which have a concept name in the same column above them, and a concept name in the same row to the left. Whether such cells hold data or not is determined by a process described later in this specification.
- 15 20 25
- 30 PROVENANCE is the source of information, including when, and in what Worksheet, the data was last changed.
- IMPLICATION is the ability to call up each Worksheet in which a selected entity is present.
- 35 CONTEXT is the additional information required to define uniquely the contents of each of the

5 data cells in a worksheet, after the worksheet has been inspected. For example if a Worksheet was to show "Total Sales" in the row headings, and "1989" in the column headings, but no further information, then the Product and Region would be the context.

10 LABEL is the information required to be attached to any item of information in order to specify that item of information uniquely. As an illustration, in a preferred example the label would usually refer to concepts by their numbers within a file of concepts, and would include the number of the entity concerned, a list of the attributes required and the numbers of the concepts that represent the instances of those attributes applicable to the data item, as well as the data type (number, alphanumeric, date, etc.).

20

In one broad aspect, the invention provides a method of setting up a spreadsheet layout having at least one concept as a row heading and at least one concept as a column heading, on a visual display unit of a computer having a database in which each item of data is stored in a separate record which includes a label in addition to the data itself, the method comprising:

30 1) naming a new worksheet and presenting it blank;

35 2) naming and positioning concepts as row headings and column headings in cells of the worksheet;

3) saving the worksheet in a usage file;

5 4) inspecting the layout to determine which
 cells should contain data to be obtained from
 the database and to determine the labels of
 any data which should be contained in each
 cell; and

10 5) searching for matches between the labels
 so determined and labels of data items in the
 database and, if found, displaying those items
 in the relevant cells of the worksheet.

15 This aspect of the invention thus allows managers to use
 spreadsheet skills to manipulate large databases stored
 separately from the headings and constants on the
 spreadsheet.

20 In a preferred form, the usage file is used to record
 the location of any of the concepts which have been defined
 in previously set up layouts. In this way, each concept may
 be used within many worksheets and may be assumed by the user
 to have the same meaning in each.

25 In another aspect, the invention provides a method as
 first defined, wherein an interface system includes rules by
 which cells in a worksheet are loaded with data and wherein
 the system is arranged to assume that the rules for loading
 any cell should be related to the row and column headings of
 the worksheet. The interface system preferably consults the
 concept definitions to construct templates or masks to
30 represent each item of data on the worksheet, including in
 the template or mask, a complete label for each item.

35 The data may be stored without regard to order or they
 may be ordered to allow faster searches of larger databases.

 In another broad aspect, the invention provides a method
 of providing a spreadsheet layout having at least one

concept as a row heading and at least one concept as a column heading, on a visual display unit of a computer having a database, the method comprising: storing each item of data in the database with the provenance thereof and, responsive to
5 selection from the computer keyboard, displaying on the visual display screen, the previous worksheet in which the data in the cell that the cursor was on when the selection was made, was last amended.

10 In addition to the date on which a data item was last changed and the worksheet within which it was changed, the provenance information may optionally include the identity of the person who last changed the data, and the time of last change. In a preferred example of the invention, provenance
15 information is stored automatically from an internal system clock, the system's knowledge of what work sheet is being used at the time of saving information, and the system's knowledge of the identity of the user requested at the start of the data session. The provenance worksheet number may
20 then be used during program execution to display data in one colour (preferably white) when the provenance is from the worksheet currently being displayed, and another colour (preferably yellow) when the provenance is from another worksheet. In this example, whenever the cursor is on a
25 particular item of data, the provenance information for that data is displayed automatically at the edge of the worksheet. The worksheet of last change recorded against each data item may then be used to allow the user, while the cursor is on a data item, to change worksheets to that in which the data was
30 last changed, setting the context of that worksheet, if required, to an appropriate context to enable the user to view and possibly edit that data item within the context of that worksheet.

35 The preferred system provides an implication facility, which is seen by the user as the converse of provenance. A request for implication will occasion a search through the

-6-

usage file for appearances of the entity represented by the current data item in other worksheets, and the presentation of a list of such worksheets for the user to select which worksheet he or she may wish to view.

In a preferred example including both broad aspects of the invention, the storage of data separate from the concept names and text of a work sheet, together with the preferred provenance information, provides the user with an unparalleled ability to view and edit a wide range of management information, and to authenticate the date on which conclusions are based.

An additional feature of the storage method is that the user can abstract a subset of one data file and merge it into another data file, thus providing a solution to the so-called "distributed database" problem of how to store and pass information between multiple files within interlinked computer systems, the files having been established by different people without reference to each other.

The preferred example of the invention will now be described with reference to the accompanying drawings, in which:

Figure 1 represents an example of some of the information in the database in stylised form;

Figure 2 shows schematically a procedure to decide what data to load into the worksheet;

Figure 3 is a flow chart showing the procedure for creating a cell table; and

Figure 5 shows a flow chart for the data search.

Each item of data is stored separately together with its label and its provenance. This is all stored in a compact form in a separate file. Another file lists all the concepts used and encoded definitions of those concepts. Figure 1 represents an example of information in the database in stylised form.

Worksheets are the principal means of editing and viewing data. Worksheets are stored separately from the data, in a compact form and as a series of records in a file of usages. Each record specifies the worksheet number, a concept number, a position on the worksheet, and a type :whether this is a concept name or an override value (see below).

Work sheets may be set up by the user as s/he pleases by positioning one or more column headings and row headings, which may have a horizontal insert within the cell in which the headings appear, and which together specify precisely some or all of the attributes of any entity the value of which the user wants to see in a cell.

While most data cells are defined by row and column headings, there are occasions when the user may want individual cells to show information which is not defined by such headings. An example might be that the user would require a single total figure at the bottom of a matrix of figures, which figure s/he would understand to be the total of the entire matrix even though there was no heading at all on the line. Or the user might want individual items of data in other places in the worksheet to which the context might still apply, but the entity would be defined specially for that cell alone. Such data items are called "override" values.

Each work sheet is held as a set of usages. Each usage is an eight byte record, consisting of 2 bytes representing the worksheet number, 2 bytes representing the concept number, 2 bytes representing the cell position (horizontal and vertical offsets from top left corner), one byte horizontal inset within each cell, and one byte usage type.

The usage type is usually FF (255 decimal) to indicate that this is the name of a concept which is either in the column headings or the left row headings. A usage type FE indicates it is in neither of these regions, it is in the middle of the work sheet. The usage type 01 indicates that this is an override value. Two special types are 43, indicating that this is a list of cosmetics, and 46 indicating that this is a list of formulae. In these two cases, a number may be found in the two bytes usually reserved for the cell position. This number refers to a record number in the file , suffixed .FML in which the lists of cosmetics and formulae may be found.

20

The database may also include other files each listing a feature or function. For example there will usually be a formulae file, which may hold both formulae and cosmetics (for the layout), and a memo file. Memos are items of text which may be written in by the user to apply to any combination of entity and attributes that s/he may care to specify. The user may optionally specify an entity and/or any single instance of each attribute. When the system loads, any memo which is relevant to the current worksheet and context is noted and displayed on demand.

30

SETTING UP A WORKSHEET

Usages are stored unordered in a file having a suffix .USA. To load the work sheet, the system first searches the usage file and extracts all those usages with the correct worksheet number. At the same time, the system

35

gauges the size of the work sheet, and allocates space in a cell index (10H , i.e. 16 decimal, bytes per cell). The index is initially cleared, then a record of each of the usages is put into the relevant index cell. The record
5 includes a flag to indicate that the cell contains a concept name and a concept number.

In order to decide what data to load into the worksheet, the system uses the following procedure (shown schematically
10 in Figure 2):

If there are no entities mentioned in the worksheet definition, the system loads and displays the worksheet, then asks the user for the name of an entity to display. Each
15 entity mentioned in the worksheet is consulted to discover the required attributes for the whole range of entities.

There is a check to discover whether the the worksheet is consistent. For this purpose the worksheet may be
20 considered to be divided into regions. Each row of column headings is a separate region. The whole area below the column headings is another separate region. If any two entities or any two instances of the same attribute are in different regions, then the worksheet is inconsistent and no
25 further settings may take place until the user has corrected the error.

If there are instances of each attribute within the worksheet definition then no further questions about that
30 attribute are asked at this stage.

If attributes are missing there are other sources whereby the system can provide defaults. If the worksheet is being set up as a result of use of the provenance or
35 implication facilities, then the item of data last accessed can be used to provide a default. There is also a table of the latest default values. The default table is updated

whenever a worksheet is quitted. If, by these means, all the context is filled in then loading will continue without asking the user to fill in context

5 If, however, there are no instances of that attribute mentioned, or those instances that are mentioned are representative, ie are not specific (see below), and no defaults are available (see below), then the system will display the worksheet and ask for values for the required
10 attributes before proceeding. This procedure is called determining the context.

Certain concepts may be defined to be representative of attributes. For example, "This Month" may be a
15 representative month. This implies that it is not a specific month but may take on the value of any given month on demand. Thus if the concepts "This month" and "Last month" are included in a worksheet, the names "This Month" and "Last month" will be included as part of the context, and the user
20 will be required to provide suitable values (e.g. February and January) before the worksheet can be loaded. The representatives are listed separately in a context window but are treated exactly like other attributes for the purpose of filling in the context.

25

The values entered (if any) will be taken to apply to any cell in the worksheet if required by that cell. However, no cell, except a cell containing override values, is considered to have data in it unless it has both row and at
30 least one column heading. If these concepts together with the context information fail to define the data in the cell entirely, the system is prepared to look at other column headings which may appear in rows which have no concepts directly above the cell in question. The next concept cell
35 to the left will be taken to be relevant if it is an example of one of the missing attributes. Similarly, the system will be prepared to look up the column that contains the row

-11-

heading (which would usually be the left-most column but need not be so) for any other concept names that may fit the required missing attribute types. If it finds such concepts above the row in question in the same column, and with an inset within the column less than the concept in the same row as the cell under consideration, then those concepts are also taken to be relevant. Figure 3 is a flow chart showing the procedure for creating a cell table.

An override value will override any value that might otherwise be expected to occupy the cell and will not necessarily require row or column headings in order to hold data. However, the cell will accept the usual rules for deciding what other concepts in the worksheet may be relevant to defining the attributes of the override cell. There is a separate table of these values stored at the end of the table of normal worksheet cells. A search is made through this table on loading data.

If a cosmetic entry has been found, the relevant cosmetics are loaded from the .FML file. These cosmetics are stored as a list of text items and their cell positions. Pointers to the individual items are inserted into the cell index.

When the user is being asked for context entries, only those attributes which are not mentioned, and those representatives that are mentioned in the worksheet definition are presented. When the user indicates that he or she has finished entry of context values, the system rechecks for completeness. If both tables are now complete, the loading may proceed.

The system must now make an internal table of those entries with which it expects to fill the work sheet when

SUBSTITUTE SHEET

searching the data. The entries in the internal table are called templates or masks of the data in the database. To build the templates, the system works methodically through each cell of the worksheet to build up a template item for
5 each cell. This process conceptually works as follows, although the practical implementation alters the order of testing in certain respects:

a) All information set by the user in the context
10 is held to be relevant to all cells and is used to start a table to relevant concepts that holds, among other items, pointers to the attribute and concept of all concepts relevant to the cell under consideration.

b) The column headings are examined to discover any
15 relevant headings. There must be at least one column heading, or it will be assumed that the cell does not contain data. If there is a heading, the attribute and concept number are placed into a relevance table. In addition, any
20 if the column heading rows that are blank are searched leftwards from the blank position, and the first concept found (if any) in those rows is also recorded, with its attribute in the relevance table.

c) The cells to the left of the current cell are
25 searched to find the nearest cell containing a concept name. If no such cell is found, it is assumed that there is no data. Otherwise the concept number and attribute number are added to the relevance table.

30

d) The column in which c) has been found is
searched upwards from the position of c) for any further
concepts that have an inset within the column less than that
of the concept c) and which are instances of attributes
35 different from those found so far (if in this context, an entity is sought, it is treated in exactly the same way as instances of attributes).

e) The list of relevant concepts is then searched to discover an entity. If no entity exists, then there is considered to be no data in the cell. If an entity is found, the cell is a data cell (ie may contain data). The definition of that entity is consulted to discover what attributes are required. The list of relevant concepts is then consulted again for each required attribute to find out if there is a complete set. If the set is complete, then a data item may be constructed.

The next available space in RAM is now used to construct a template of the data that will fill that cell. This template is identical to the data record that may be found, except that it contains no data (the data field is left blank as is the provenance field). The object of this template is to use it to compare byte for byte with data items that are found to do a final check for suitability when loading data. It will also be used as the framework within which data for each cell is edited in RAM, and the whole record will be stored away if needed at the point of saving.

Once the template is constructed, a flag is inserted into the cell index to indicate that the cell contains data, and a pointer to the data item is also set in the cell index.

The practical implementation of these results differs from the conceptual process in two respects. First, the system scans the worksheet starting at the top left corner and, working rightwards, down each column in turn. By this means it is not necessary to rebuild the relevance table for each cell, but rather to keep the same table continuously updated, simply overwriting irrelevant data as further entries are found in the column scans, and even keeping some of the table when a new column is scanned. Second, for each cell, the system commences making a template before it knows that it has all the relevant attributes, so necessitating

only one pass through the attribute check. If attributes are missing, the construction of the template for that particular cell is aborted, and the space freed for reuse.

5 DATA SEARCH

A list or table of relevant concepts is made up of all the concepts used in the worksheet being set up (including the context concepts) in numerical order. This list is
10 padded out with FFFFH (65535 decimal) numbers at the end to make its length an exact power of 2 (4, 8, 16, 32 etc.), so that it may be subject to a binary chop technique search. Each concept (eg entity or attribute) in the label of each data item in the database is then checked for a match
15 with the concepts in the list of relevant concepts above. If any fail then the item is rejected.

If the item passes the test above, then each column of the worksheet is checked for compatibility by investigating
20 the column headings. If any column heading is found to be incompatible with the particular item of data, then the columns are skipped until that column heading no longer applies. If all column headings fit the data item, then the pointers to the templates from the data cell index in that
25 column are used to check each template in the column against the data item byte for byte. If the entity and attribute information match, then the item is considered to be found.

Usually if an item is found, then the provenance
30 information and the data information will simply be transferred from the data file to the template. If, however, there is found to be a mismatch of data types, then the data item is converted to the new form before being stored in the template. (Both the data item and the template contain bytes
35 to indicate the data type: this byte is set when setting the template, on the basis of the definition of the concepts involved. If the concept definitions are changed, e.g.

numbers are changed to alphanumerics or the number of decimal places is changed then the data type may have changed between that expected and that found). Figure 5 shows a flow chart for the data search. Any single item of data may occur more
5 than once in a worksheet, so the search through the templates continues even after a match has been found.

QUITTING A WORK SHEET

10 When quitting a worksheet, the system will automatically save any specific attribute from the context to use as a default should that attribute be mentioned in a further worksheet, whether or not it is in the next worksheet accessed. If the change of worksheet is occasioned by a
15 provenance or implication command, the full template of the cell on which the cursor is sitting is extracted from the list of templates to be put in a special location to be used in the attribute default procedure. At the relevant stage of loading the new work sheet, this template is used to set any
20 un-ascribed attributes, or un-ascribed representative concepts, as may be required to ensure that the particular item of data extracted has the best chance of being displayed on the worksheet accessed. Once the worksheet has been loaded, the system then does a search through the templates
25 to discover the match with the extracted data, from which it may set the cursor correctly on the same data, and when it finds the correct data template, the system transfers in the data from the original worksheet, whether or not any alteration was saved when exiting the old worksheet.

30

35

CLAIMS

1. A method of setting up a spreadsheet layout having at least one concept as a row heading and at least one concept as a column heading, on a visual display unit of a computer having a database in which each item of data is stored in a separate record which includes a label in addition to the data itself, the method comprising:
 - a) naming a new worksheet and presenting it blank;
 - b) naming and positioning concepts as row headings and column headings in cells of the worksheet;
 - c) saving the worksheet in a usage file;
 - d) inspecting the layout to determine which cells should contain data to be obtained from the database and to determine the labels of any data which should be contained in each cell; and
 - e) searching for matches between the labels so determined and labels of data items in the database and, if found, displaying those items in the relevant cells of the worksheet.
2. A method as claimed in Claim 1, wherein the usage file is used to record the location of any of the concepts which

have been defined in previously set up layouts.

3. A method as claimed in Claim 1 or 2, wherein an interface system includes rules by which cells in a worksheet are loaded with data and wherein the system is arranged to assume that the rules for loading any cell should be related to the row and column headings of the worksheet.
4. A method is claimed in Claim 3 wherein the interface system consults the concept definitions to construct templates or masks to represent each item of data on the worksheet, including in the template or mask, a complete label for each item.
5. A method as claimed in any preceding Claim including storing each item of data in the database with the provenance thereof and, responsive to selection from the computer keyboard, displaying on the visual display screen, the previous worksheet in which the data in the cell that the cursor was on when the selection was made, was last amended.
6. A method as claimed in Claim 5, wherein the provenance information includes the identity of the person who last changed the data, and the time of last change.
7. A method as claimed in Claim 5 or 6 wherein provenance

18

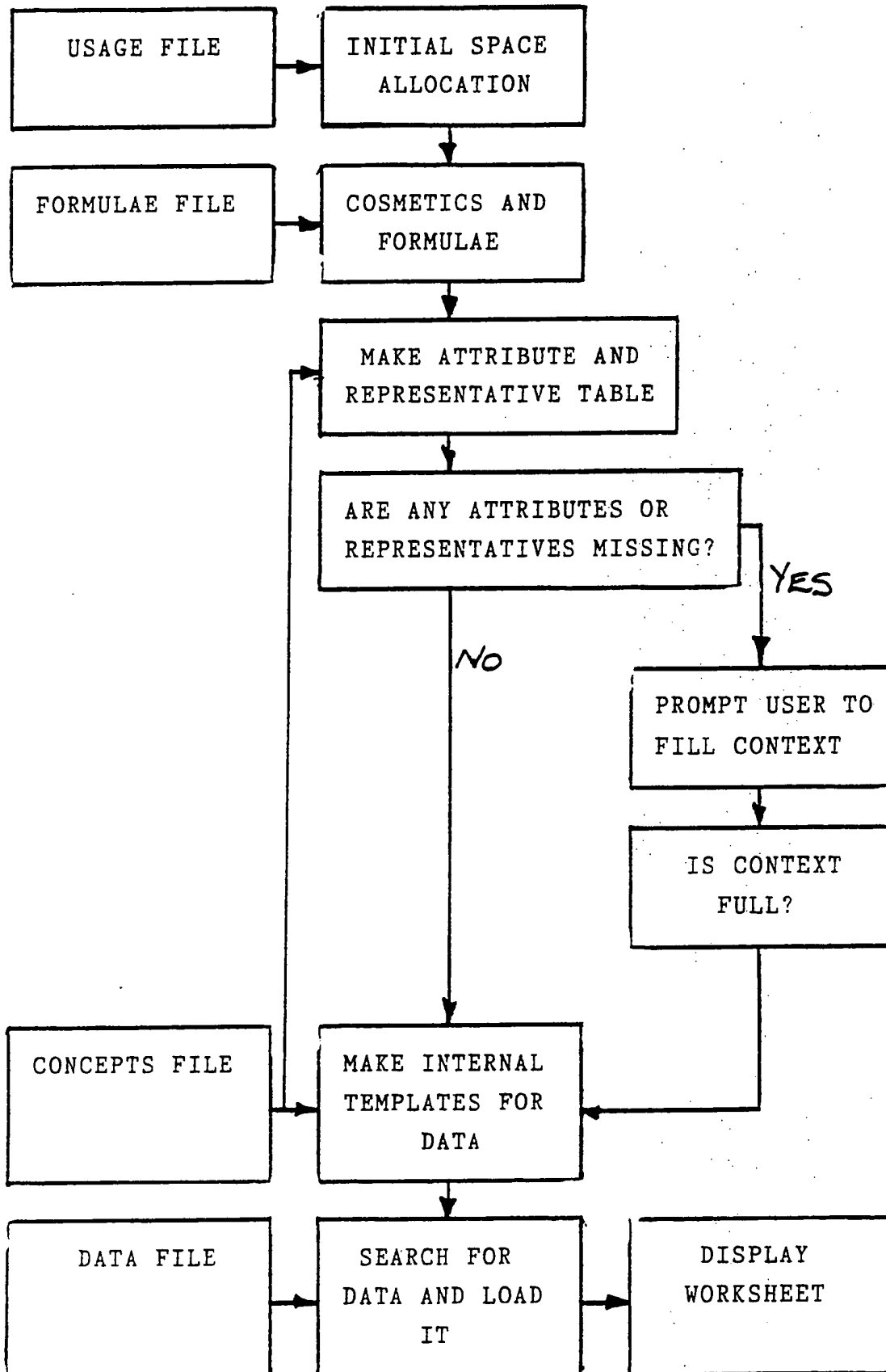
information is stored automatically from an internal system clock, the system's knowledge of what worksheet is being used at the time of saving information, and the system's knowledge of the identity of the user requested at the start of the data session.

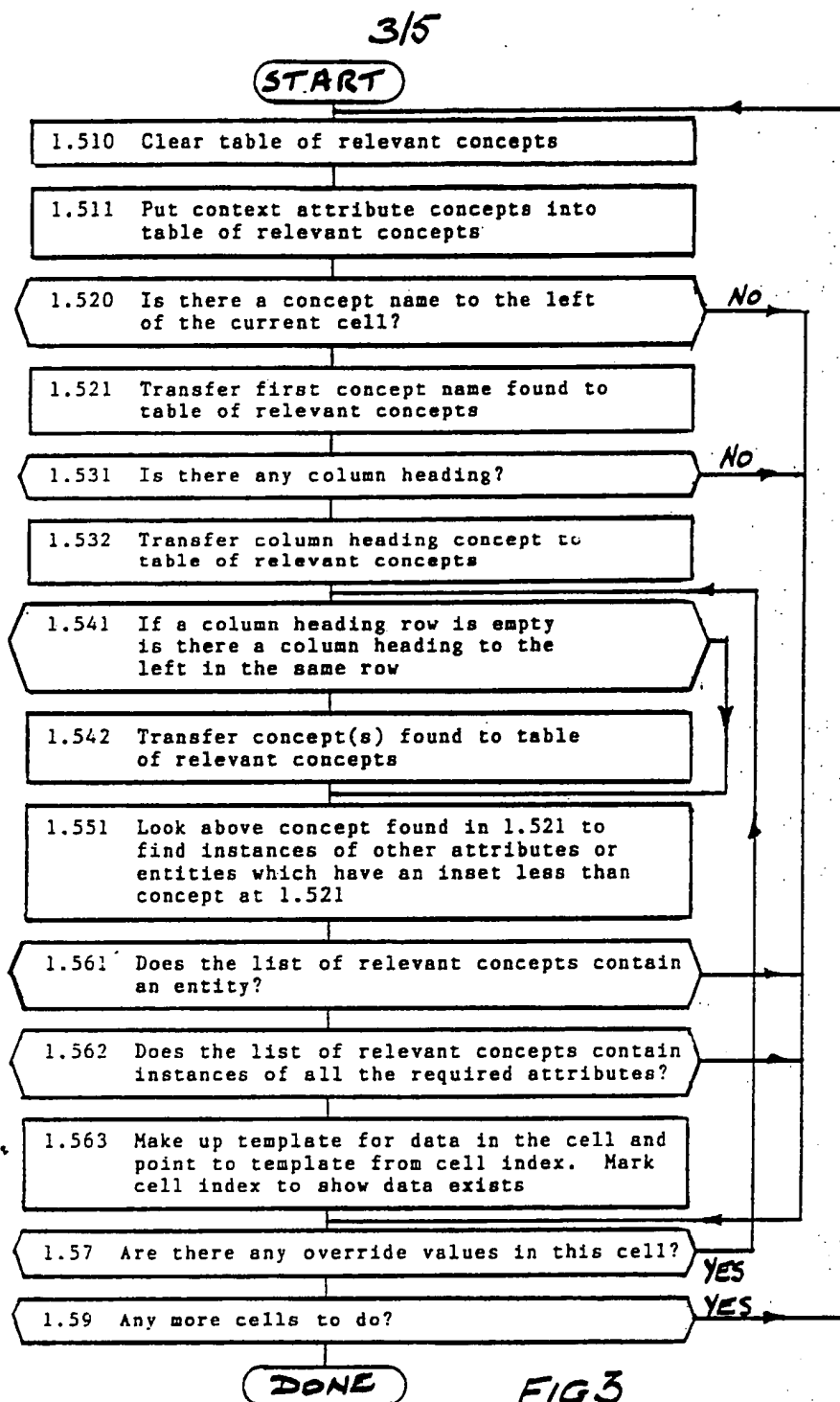
8. A method as claimed in Claim 5, 6 or 7, including:
responsive to a request for implication, conducting a search through the usage file for appearances of the entity represented by the current data item in other worksheets, and presentating a list of such worksheets for the user to select which worksheet s/he may wish to view.

1/5

FIGURE 1

<u>DATA</u> <u>FILE</u>				
VALUE	LABEL		PROVENANCE	
	ENTITY	ATTRIBUTES		(WKST = Worksheet)
1200	Number sold	<u>Model</u>	<u>Region</u>	<u>Salesman</u>
		GL	S East	Jones
1800	Number sold	<u>Model</u>	<u>Region</u>	<u>Salesman</u>
		GS	all	Jan
1700	Stock held	<u>Model</u>	<u>Region</u>	<u>Period</u>
		all	all	1988
12	Stock held	<u>Model</u>	<u>Region</u>	<u>Period</u>
		GS	East	Feb
95p	Price	<u>Item</u>	<u>Shop</u>	<u>Date</u>
		pencil	Harrods	Jan 19
£1.23	Price	<u>Item</u>	<u>Shop</u>	<u>Date</u>
		pad	WHSmith	Jan 19
				WKST 124, Jan 19, 10.11, Pat
				WKST 92, Aug 17, 12.00, Fred
				WKST 87, Aug 19, 17.00, Jim
				WKST 21, Jan 05, 08.00, Ron
				WKST 37, March 1, 09.00, Dick
				WKST 123, Jan 19, 10.10, Ann

FIGURE 2



**Page not transmitted by the receiving
Office to the International Bureau at
the time of publication**

FIG 5